Plant Molecular Biology **24**: 835–851, 1994.

Molecular analysis of the aspartate kinase-homoserine dehydrogenase gene from *Arabidopsis thaliana*

Marc Ghislain [1], Valérie Frankard [1], Dirk Vandenbossche [1], Benjamin F. Matthews [2] and Michel Jacobs [1]

[1] *Laboratory for Plant Genetics, Vrije Universiteit Brussel, Paardenstraat 65, B-1640 Sint-Genesius Rode, Belgium;* [2] *Plant Molecular Biology Laboratory, US Department of Agriculture, Agricultural Research Service, Beltsville, MD 20705, USA*

Received 3 May 1993; accepted in revised form 25 November 1993

Key words: amino acid biosynthesis, aspartate kinase homoserine dehydrogenase, gene structure, *Arabidopsis thaliana*

Abstract

The gene encoding *Arabidopsis thaliana* aspartate kinase (ATP:*L*-aspartate 4-phosphotransferase, EC 2.7.2.4) was isolated from genomic DNA libraries using the carrot *ak-hsdh* gene as the hybridizing probe. Two genomic libraries from different *A. thaliana* races were screened independently with the *ak* probe and the *hsdh* probe. Nucleotide sequences of the *A. thaliana* overlapping clones were determined and encompassed 2 kb upstream of the coding region and 300 bp downstream. The corresponding cDNA was isolated from a cDNA library made from poly(A)$^+$-mRNA extracted from cell suspension cultures. Sequence comparison between the *Arabidopsis* gene product and an AK-HSDH bifunctional enzyme from carrot and from the *Escherichia coli thrA* and *metL* genes shows 80%, 37.5% and 31.4% amino acid sequence identity, respectively. The *A. thaliana ak-hsdh* gene is proposed to be the plant *thrA* homologue coding for the AK isozyme feedback inhibited by threonine. The gene is present in *A. thaliana* in single copy and functional as evidenced by hybridization analyses.

The apoprotein-coding region is interrupted by 15 introns ranging from 78 to 134 bp. An upstream chloroplast-targeting sequence with low sequence similarity with the carrot transit peptide was identified. A signal sequence is proposed starting from a functional ATG initiation codon to the first exon of the apoprotein. Two additional introns were identified: one in the 5′ non-coding leader sequence and the other in the putative chloroplast targeting sequence. 5′ sequence analysis revealed the presence of several possible promoter elements as well as conserved regulatory motifs. Among these, an *Opaque2* and a yeast GCN4-like recognition element might be relevant for such a gene coding for an enzyme limiting the carbon-flux entry to the biosynthesis of several essential amino acids. 3′ sequence analysis showed the occurrence of two polyadenylation signals upstream of the polyadenylation site.

This work is the first report of the molecular cloning of a plant *ak-hsdh* genomic sequence. It describes a promoter element that may bring new insights to the regulation of the biosynthesis of the aspartate family of amino acids.

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers X71363 (*A. thaliana* (Landsberg *erecta*) *ak-hsdh* gene) and X71364 (*A. thaliana* (Columbia) *ak-hsdh* gene).

Abbreviations: AK, aspartate kinase; HSDH, homoserine dehydrogenase; ID, intermediate domain; Tp, transit peptide.

Introduction

Aspartate is the precursor of the essential amino acids lysine, threonine, methionine and isoleucine. In bacteria, aspartate kinase (ATP:*L*-aspartate 4-phosphotransferase, EC 2.7.2.4) catalyses the first step of this pathway, the ATP-dependent addition of phosphate to aspartate. This enzyme is present in all Enterobacteriaceae examined as three distinct isozymes: AK I-HSDH I encoded by the *thrA* gene, AK II-HSDH II encoded by the *metL* gene and AK III encoded by the *lysC* gene, reviewed by Cohen and Saint-Giron [1]. Various mechanisms of regulation have been found for this enzyme: activation, repression, growth condition-dependent expression of the AK isozymes and feedback inhibition of the isozymes by different end-products.

In higher plants, aspartate-derived amino acid biosynthesis occurs through a similar pathway. It is controlled mainly by feedback inhibition of several branch point enzymes by end products of the pathway [2]. These key enzymes play an important role in determining the amount of free amino acids such as lysine and threonine which are respectively the first- and second-limiting amino acids in diets based on cereals for monogastric animals [3]. Disruption of these regulatory loops in higher plants led to the overproduction of lysine or threonine [4]. To date only feedback regulation has been clearly documented in higher plants. At least two AK isozymes have been found in higher plants: a first isozyme is inhibited by lysine and synergistically by lysine plus *S*-adenosyl methionine, and a second by threonine. There is no clear evidence in higher plants for the presence of an insensitive isozyme as there is in *E. coli*. Modification of the feedback inhibition of the lysine-sensitive aspartate kinase isozyme (*Lys*-AK) has been shown to result in a higher level of threonine in the free amino acid fraction in barley [5], carrot [6], maize [7], *Nicotiana sylvestris* [8] and *Arabidopsis thaliana* [9]. *Lys*-AK has been partially purified from barley [10], carrot [11], maize [12] and *N. sylvestris* [8] and shown in the two monocots to be under the control of two unlinked loci [5, 7].

Molecular analysis of aspartate kinase in the model plant *A. thaliana* aims at the elucidation of the isozyme identification and expression during the plant development as well as at the isolation of the *lys-ak* allele, desensitized to lysine inhibition in the *A. thaliana* mutant 5 FALT 40/6/1 [9] and in *N. sylvestris* mutant RLT70 [8]. Recently, one cDNA coding for one of the carrot AK isozymes has been cloned and its translation product was shown to be a bifunctional protein with both aspartate kinase and homoserine dehydrogenase activities [13, 14]. We have used the carrot cDNA as a probe to isolate the corresponding gene(s) in *Arabidopsis thaliana* and to analyse its structure.

Materials and methods

Carrot ak-hsdh *cDNA as a probe*

The carrot cDNA is divided into three functional domains: AK for the aspartate kinase activity, ID for the intermediate domain, HSDH for the homoserine dehydrogenase activity [14]. The *Eco* RI-*Eco* RV 1.2 kb fragment and the *Eco* RI 1.1 kb fragment of the carrot *ak-hsdh* cDNA covering respectively the whole AK region and the HSDH region will be further referred to as the *ak* probe and the *hsdh* probe.

Plant gene source

Arabidopsis thaliana race (Landsberg Erecta) was used for Southern blot analysis. Two genomic banks were kindly provided by Dr H. Goodman, Boston, USA. The *A. thaliana* race Columbia λEMBL3 and *A. thaliana* race Landsberg Erecta λFix genomic banks were screened respectively with the *ak* probe and the *hsdh* probe.

DNA manipulations

Recombinant techniques were used for cloning, screening and hybridization [15]. Genomic DNA fragments were subcloned in pUC18 for DNA sequencing and transformed [16] into *E. coli* XL1-Blue cells (Stratagene).

DNA hybridizations

Whole plants were crushed in liquid nitrogen and total DNA was extracted following the rapid procedure of Dellaporta *et al.* [17]. Restriction enzymes were used as described by the manufacturer (Boehringer) at 2 units per μg in a 4 h reaction. Southern blots of genomic DNA contained 10 μg of DNA per lane with the DNA fragments separated on a 0.8% agarose gel. Electrophoresis and transfer to Hybond-N (Amersham) membrane were done according to standard procedures with a vacuum blotting system (Pharmacia-LKB). Genomic DNA gel blots were prehybridized at least 2 h and hybridized overnight at 42 °C in the following buffer: 5 × SSC, 30% (v/v) desionized formamide, 0.1% (w/v) SDS, 10 × Denhardt's solution, 100 μg/ml hydroxylated salmon sperm DNA, 10 mM Tris-HCl pH 7.5, 1 mM EDTA. DNA probes were prepared from 25 ng of agarose gel-eluted DNA and labeled with the T_7 DNA polymerase random-primer labelling kit (Pharmacia). Membranes were washed in 1 × SSC + 0.1% (w/v) SDS once 20 min at room temperature and twice at 42 °C for 30 min. Plaque hybridization was performed identically.

Screening of a cDNA library

Screening of a λ UNI ZAP II cDNA library constructed from poly(A)$^+$-mRNA extracted from cell suspension cultures (kindly provided by Dr Trezzini, Max Planck Institut, Germany) was done using the *Spe* I restriction fragment (800 bp) from plasmid pATAK2. From ca. 110000 plaques, six clones were isolated among which only one seemed to be full-length. All clones however displayed similar restriction digestion patterns, differing only in their total length.

RNA extraction and hybridization

Poly(A)$^+$-mRNA was extracted from rapidly growing cell suspension cultures from callus explant of *A. thaliana* race Columbia (kindly provided by Prof. L. Willmitzer, University of Berlin, Germany) using a batch method based on oligodT affinity chromatography (Quickprep micro mRNA Purification Kit, Pharmacia). Northern blotting was performed according to Fourney *et al.* [18].

DNA sequencing

The nucleotide sequence was determined on both strands of denatured plasmid DNA using the dideoxy chain-termination method [19] with the Sequenase version 2.0 sequencing kit (United States Biochemical Corp.).

Promoter-GUS fusion

The construction of the chimeric gene was initiated by PCR amplification of the pATAK2 insert using the reverse primer and the 23-mer oligonucleotide CATGGCGTAACTCAGTCAAACAC complementary to the sequence beginning at position 2045. The 350 bp amplification product was digested by *Hin*d III and subcloned into pUC18 digested with *Hin*d III and *Sma* I. The nucleotide sequence was verified to avoid possible PCR mistakes. The *Hin*d III-*Nco* I fragment was finally subcloned in the corresponding sites of pHW8 (Dr J Botterman, PGS, Belgium). This construct is referred to as pEPAK3. pEPAK4 was obtained by removing the *Bam* HI-*Bgl* II fragment from pEPAK3, leaving a 270 bp fragment. The expression study was conducted with this construct.

Transient gene expression study

Protoplasts from cell suspensions of *Nicotiana plumbaginifolia* and the transient gene expression experiment were performed according to Bilang and Schnorf [20] with a modified enzyme solution (Drieselase 0.4%, Cellulase 0.5%, Macerozyme 0.5%). GUS enzymatic activity was measured essentially as described by Jefferson [21].

Results

ak *and* hsdh *DNA sequences are clustered within a single locus*

Genomic DNA of *A. thaliana* race Landsberg *erecta* was analysed on Southern blots to determine the complexity of the hybridization pattern with the *ak* probe (Fig. 1A) and *hsdh* probe (Fig. 1B). DNA hybridization was first performed under relaxed hybridization conditions with the *ak* probe. After 3 days of autoradiography, the membrane was washed and rehybridized with the *hsdh* probe. Comparison of the two autoradiographies shows that the *Eco* RI and *Hin*d III digestions give different hybridization signals whereas with the *Bam* HI digestion, the signal is exactly at the same position. No other hybridization signals could be detected by Southern blot analysis although relaxed hybridization conditions were used. This was our first evidence that *ak* and *hsdh* were probably also clustered in a single locus in *Arabidopsis*.

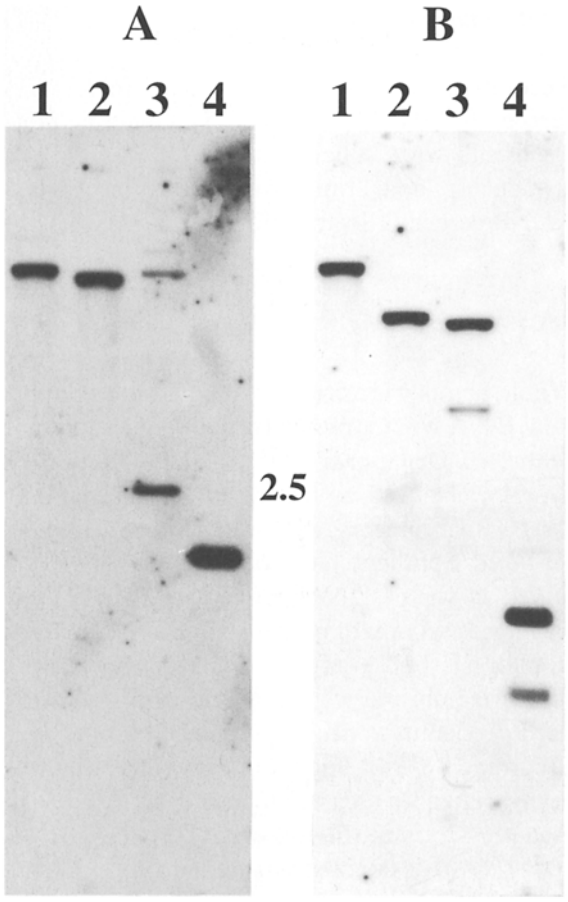


Fig. 1. Hybridization pattern between *Arabidopsis thaliana* total DNA and the carrot gene. 10 μg of *A. thaliana* (race Lansdberg *erecta*) genomic DNA restricted by *Bam* HI (lane 1), *Eco* RI (lane 2), *Hin*d III (lane 3). Lane 4 is 100 pg of *Eco* RI-*Eco* RV restricted *ak-hsdh* cDNA of carrot. The filter was successively hybridized with the *ak* probe (A) and the *hsdh* probe (B).

ak-hsdh *nucleotide sequence and gene structure*

The carrot *ak* probe was used to screen a genomic bank, λEMBL3 *A. thaliana* race Columbia (Fig. 2). One positive clone, λE19, was further characterized and shown to contain the AK encoding region and a 10 kb fragment upstream. However the 3′ end of the gene was not present on this clone. The 2.8 kb, 2.7 kb, 5.6 kb *Hin*d III fragments and the 5 kb *Eco* RI-*Sal* I fragment of λE19 were subcloned into pUC 18 and designated pATAK1, 2, 3 and 4. Nucleotide sequencing showed that they contained homologous regions to the carrot *ak* and *id* regions (Figs. 3 and 5a). Screening with the *hsdh* probe did not allow the isolation of a new clone. This is probably due to the different amplifications performed on that genomic bank. Therefore another genomic bank, λFIX *A. thaliana* race Landsberg *erecta*, was screened with the *hsdh* probe and one clone among the 5 positive clones isolated, λF211 was shown to overlap the λE19 and to contain the 3′ end of the gene. The 1.8 kb *Eco* RI-*Sal* I fragment and the 2.2 kb *Eco* RI fragment were subcloned

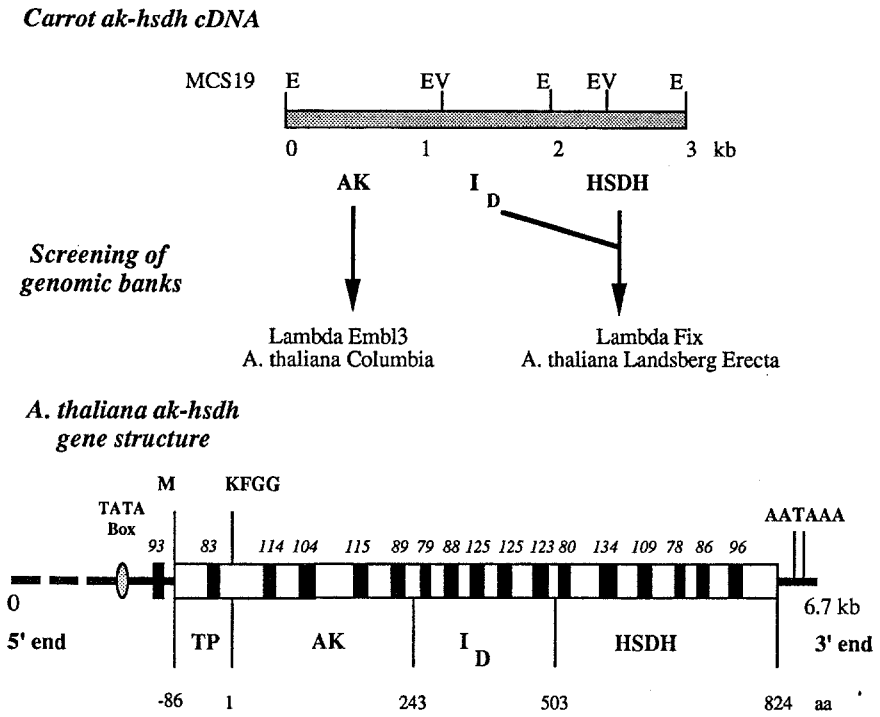
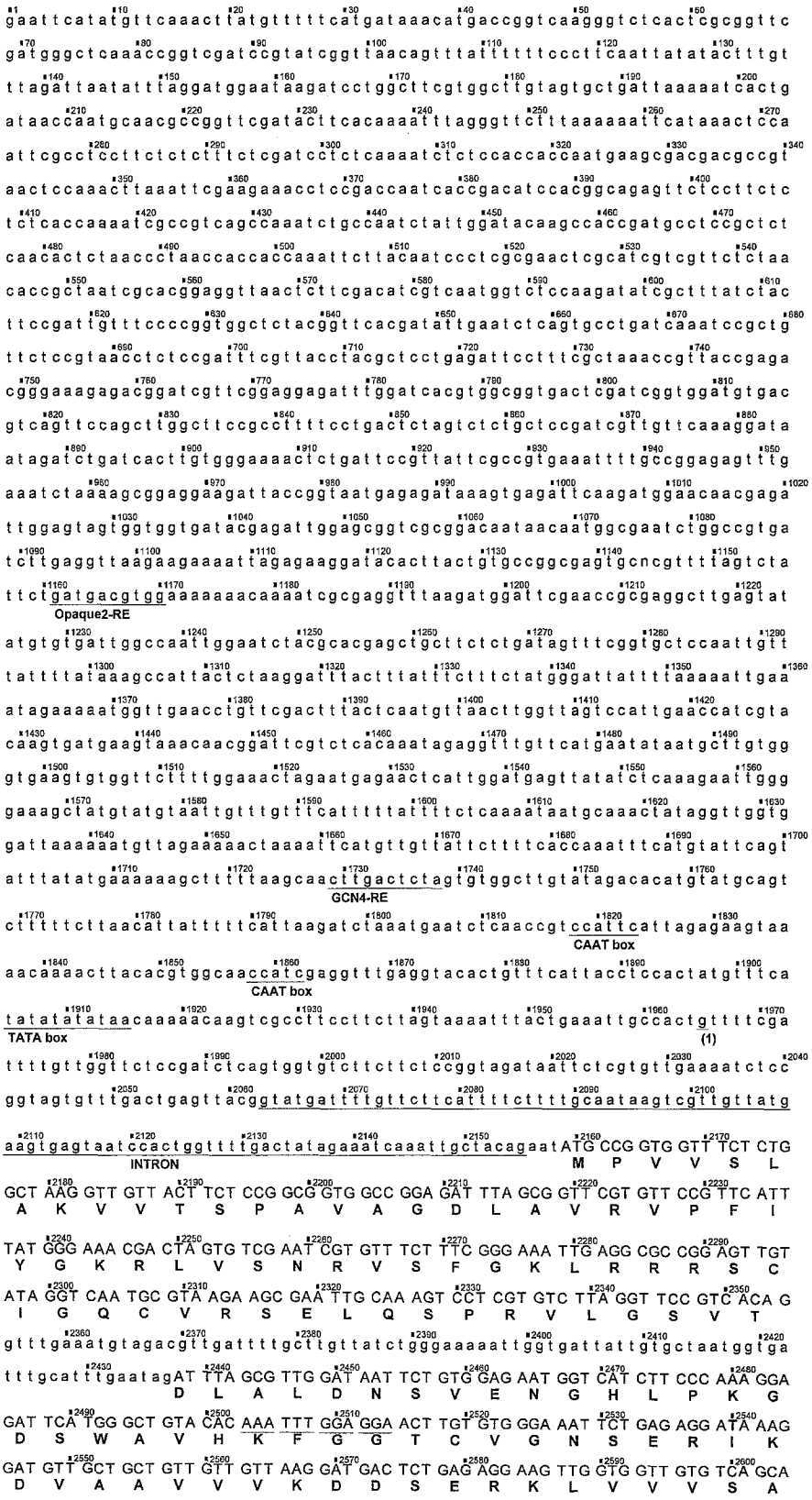
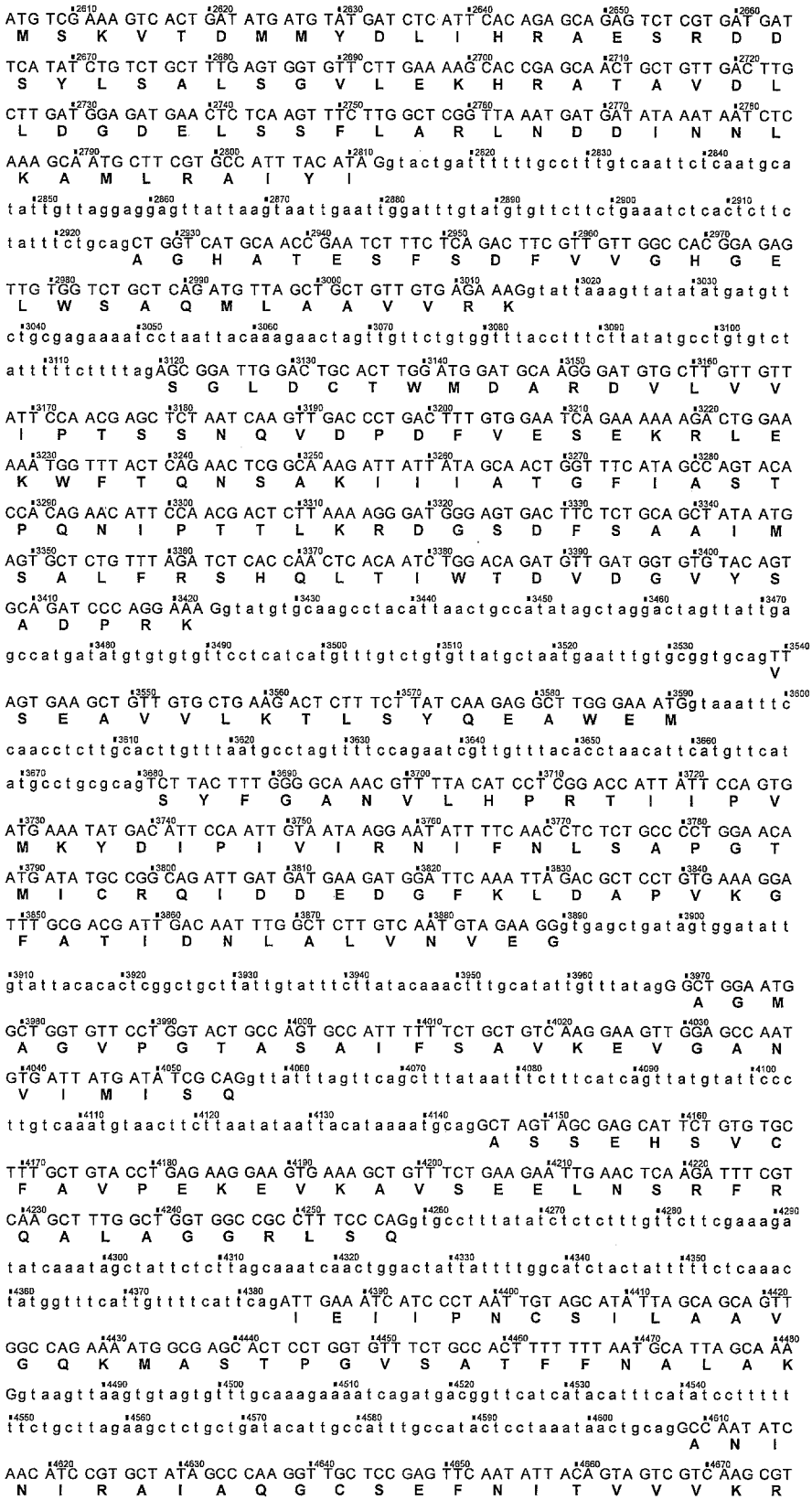


Fig. 2. ak-hsdh gene structure and cloning strategies. Carrot *ak-hsdh* restriction map is presented with the following abbreviations: MCS 19, multiple cloning site of vector pUC19; E, *Eco* RI: EV, *Eco* RV. White boxes are the exons of the *Arabidopsis thaliana ak-hsdh*-coding region. Length of introns (black boxes) is indicated above each of them. M represents the initiation codon while KFGG is the 5′ end conserved AK amino acid sequence. TP, AK, I_D, HSDH are the functional domains, respectively the transit peptide, aspartate kinase, intermediate domain and homoserine dehydrogenase. Overlapping regions of the gene are from pATAK2 and pATHD.

and designated pATHD1 and 2. Nucleotide sequencing revealed that λF211 contained a 1.4 kb overlapping region with the clone pATAK4, the end of the gene and a 10 kb fragment downstream. Restriction analysis of the genomic fragments revealed a physical map in agreement with the Southern blotting analysis. The nucleotide sequence of the *Arabidopsis ak-hsdh* gene resulting from the sequencing of part of λE19 and λF211 phages is presented in Fig. 3.

The position of the exons were determined by homology comparisons with the carrot and *A. thaliana* cDNAs and using the GT...AG boundaries rule of the introns. Translation of the exons of the *Arabidopsis* gene shows that it also encodes a bifunctional AK-HSDH protein. The NH_2 end of AK-HSDH, a chloroplast-localized enzyme, is expected to begin with a transit peptide (Tp) sequence as observed with the carrot enzyme. The localization of the putative Tp sequence of the *A. thaliana ak-hsdh* gene derives from the nucleotide sequence comparison with a cDNA which was isolated using as probe the 5′ end of the gene (the 800 bp *Spe* I fragment of pATAK2). This cDNA which begins at position 1965 and ends at position 6645 displays one long open reading frame of 2736 bp starting with a first ATG at position 2159 bp. One small intron was found in the 5′ non-coding region of the gene, and another in the putative Tp sequence. Assuming the identified ATG is the *ak-hsdh* initiation codon, we propose a Tp sequence which best fits the amino acid composition of the chloroplast T_p sequences [22], the conserved sequences of the 5′ and 3′ junctions for plant introns according to Brown [23] and the sequence similarities with the other plant *ak-hsdh* sequence [14] (Figs. 3 and 5a). The nucleotide sequence of the transit peptide is also


Opaque2-RE
GCN4-RE
CAAT box
CAAT box
TATA box
(1)
INTRON
M P V V S L
A K V V T S P A V A G D L A V R V P F I
Y G K R L V S N R V S F G K L R R R S C
I G Q C V R S E L Q S P R V L G S V T
D L A L D N S V E N G H L P K G
D S W A V H K F G G T C V G N S E R I K
D V A A V V V K D D S E R K L V V V S A




polyA signal
polyA signal
(2)

taken into account since it appears to be more conserved than the amino acid sequence as reported by Kaneko *et al.* [24]. The Tp sequence most probably ends before the **KFGG** motif, thus being of around 80 amino acids long. Two evidences support this assumption: the amino acid sequence identities drop down abruptly upstream of this sequence, and this motif has been found at the amino end of all the AK characterized to date including *Escherichia coli* [25–27], *Serratia marcescens* [28], *Bacillus subtilis* [29], *Corynebacterium glutamicum* [30, 31], *Saccharomyces cerevisiae* [32] and higher plants [14].

The AK-HSDH apoprotein coding sequence is interrupted by 15 introns. Together with the leader and Tp sequences, a total of 17 introns are identified. They are small, ranging between 78 and 134 nucleotides in length, a common feature of *Arabidopsis* genes. Their consensus sequence fits very well with the plant splice junction consensus sequence [23]. Using the intron classification of Hanley and Schuler [33], based on the purine or pyrimidine content of the sequences upstream of the 3′ splice site, 8 introns can be classified as class I (pyrimidine-rich), 3 as class III (purine-rich) and 4 as class II (mixed class). Interestingly, this distribution fits best with that of mammalian genes rather than that of dicot plant genes. Indeed, this gene has an unusual proportion of pyrimidine-rich introns.

This gene is functional in *A. thaliana* as shown by northern blotting experiments (Fig. 4). Indeed, a 3.2 kb transcript was detected in the poly(A)$^+$-mRNA fraction extracted from cell suspension cultures.

The amino acid sequence deduced from the *Arabidopsis ak-hsdh* gene is compared to the other bifunctional enzymes of carrot and *E. coli* in Fig. 5a. Analysis of the structure of the putative gene AK-HSDH was done taking for convenience the first residue of the KFGG conserved motif as position 1. The gene product is made of 824 residues and has a calculated molecular mass

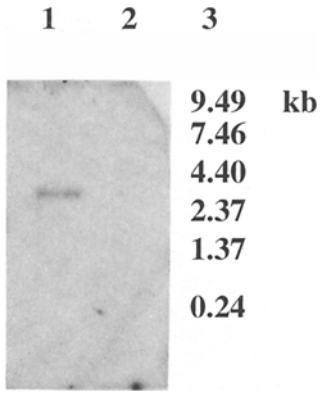


Fig. 4. Northern blot with the *Arabidopsis thaliana ak-hsdh* gene. Lanes 1 and 2 correspond to 3 μg RNA from cell suspension cultures respectively the poly(A)$^+$-mRNA and the total fractions. In lane 3 are the molecular weight markers.

of 90065 Da. The so-called KFGG motif is the core sequence of a highly conserved stretch present nearly immediately at the beginning of the protein. Two putative functional amino acid conserved sequences are also boxed in Fig. 5a: the D-P-R sequence, which is highly conserved among all the AKs and the G-X-G-X-X-G sequence which is most probably the NADPH-binding domain of homoserine dehydrogenase [28, 34]. The D-P-R sequence which is in the first half of the AK conserved sequence (Fig. 5b, top), is likely to determine the kinase activity since it is significantly homologous to the putative kinase domain of the δ-glutamyl kinase from the *E. coli proB* gene [35] and *Vigna aconitifolia* P5CS gene [36] (Fig. 5b, bottom). Codon usage of the *ak-hsdh* gene indicates a strong preference in the third position for A or T, since 64% of codons are XXA/T, a rather strong bias. This value is in agreement with the observed distribution of codon frequency for dicot genes [37]. The *Arabidopsis* apoprotein shows 80.3% amino acid identity to the carrot sequence and is shorter by 2 amino acids (position 313). Comparison of the plant and *E. coli* bifunctional protein sequences reveals that the homology is not evenly distributed along the

Fig. 3. Nucleotide and deduced amino acid sequences of the *Arabidopsis thaliana ak-hsdh* gene. Promoter elements as well as polyadenylation signals are underlined. RE stands for regulatory element, (1) and (2) mark the cDNA ends.

[a] Amino acid sequence

```
Dc                                                -95  SLSSAISP
At                                                -87        MP

At   VVSLAKVVTSPAVAGDLAVRVPFIYGKRLVSNRVSFGKLRRRSCIGQCVR
Dc   SSYAAIAAAYSARTPIFNKKKTAAVLSPLSLFHQSPSLSKTGIPLHRGRK

                                                      1
At   SELQSPRVLGSVTDLA  LDNSVENGHLPKGDSWAVH KFGG TCVGNSERI 13
Dc   ESSSKFYIAA---TAVPS--D---KV---R-AM-SI---------S---- 13
EcI                 -4                  MR-L-----SLA-A--F 13
EcII               -15     MSVIAQAGAKGRQL-----SSLADVKCY 13

At   KDVAAVVVKDDSERKLV VVSAMSKVTDMMYDLIHRAESRDDSYLSALSG 62
Dc   RN--EI--E-------- ----------------YK-Q------E---DA 62
EcI  LR--DILESNARQGQVAT-L--PA-I-NHLVAM-EKTI-GQ-    --PN 59
EcII LR--GIMAEYSQPDDMM ----AGST-NRLISWL  KL-QT-R --- HQ 58

At   VLEKHRATAVDLLDGDE       LSSFLARLNDDINNLKAMLRAIYIAG 105
Dc   -M---KL--F------D       -AR--T--QH-V-------------- 105
EcI  ISDAE-IF- E--T-LAAAQPGFP-AQLKTFVDQEFAQI-HV-HG-SLL- 108
EcII -QQTL-RYQC--IS-LLPAEEA   D-LISAFVS-LER-A-L-DS      100

                    Aspartate Kinase
At   HATESFSDFVVGHGELWSAQMLAAVVRKSGLDCTWMDARDVLVVIPTSSN 155
Dc   --------------------L-SF-I--N-G--N---T------N-AG-- 155
EcI  QCPD-INAALICR--KM-IAIM-G-LEAR-        HN-T-ID-VEKL 150
EcII GINDAVYAE------V---RLMS--LNQQ--PAA-L---EF-RAERAAQP 150

At   QVDPD  F VESEKRLE   KWFTQNSA   KIIIATGFIASTPQNIPTT 196
Dc   -----  Y L-------   ---SS-QC   QT-V--------------- 196
EcI  LAVGH  Y L--TVDIAESTRRIAASRIPADHMVLMA--T-GNEKGELVV 197
EcII ---EGLSYPLLQQLLVQHPG-RL           VV----SRNNAGETVL 189

                                                    243
At   LKRDGSDFSAAIMSALFRSHQLTIWTDVDGVYSA DPR KVSEAVVLKTLSY 246
Dc   -------------G--L-AG-V------N---------------------- 246
EcI  -G-N---Y---VLA-CL-ADCCE-----N---TC---Q-PD-RL--SM-- 247
EcII -G-N---Y--TQIG--AGVSRV---S--A----------KD-CL-PL-RL 239

At   QEAWEMSYFGANVLHPRTIIPVMKYDIPIVIRNIFNLSAPGTMICRQI   294
Dc   ----------------------R----------------------ESVG 296
EcI  ---M-L-----K-------T-IAQFQ--CL-K-TG-PQ----L-   GAS 294
EcII D--S-LARLA-P---A--LQ--SGSE-DLQL-CSYTPD        -GST 281

At   DDEDGFKLDAPVKGFATIDNLALVNVEGAGMAGVPGTASAIFSAVKEVGA 344
Dc   ET---L--ESH-----------I----T------------G---D--- 346
EcI  R---EL    ----ISNLN-M-MFS-S-P--K-MV-M-ARV-A-MSRARI 340
EcII RI-RVLASGTGARIVTSH-DVC-IEFQVPASQDFKLGHKE-DQIL-RAQV 331

                    Intermediary Domain
At   NVIMISQASSEHSVCFAVPEKEVKAVSEELNSRFRQALAGGRLSQIEIIP 394
Dc   -------------I------S-----AKA-EA------DA-----VA--- 396
EcI  S-VL-T-S---Y-IS-C--QSDCVRAERAMLEE-YLE-KE-L-EPLAVAE 390
EcII RPLAVGVHNDRQLLQ-C YTS--ADSALKILDE      A-LPGELRLRQ 374

At   NCSILAAVGQKMASTPGVSATFFNALAKANINIRAIAQGCSEFNITVVVK 444
Dc   ------T--------------L----------V-------T-Y-----LS 446
EcI  RLA-ISV--DGLRTLR-I--K--A---R-----V------S--RS-S---N 440
EcII GLALV-M--AGVTRN-LHCHR-WQQ-KGQPVEFTW  -SDDGISLVA-LR 422

At   REDCIRALRAVHSRFYLSRTTLAVGII GP GLI GGTLLDQIRDQAAVLKEE 494
Dc   ----V---K------------I----V------A-----L-----I---N 496
EcI  ND-ATTGV-VT-QMLFNTDQVIE-FV--V-GV--A--E-LKR-QSW-- N 489
EcII TGPTESLIQGL-QSVFRAEKRIGLVLF-K-N--SRW-ELFAREQST-SAR 472
```

```
                503
At   FKIDLRVIGITGSSKMLMSESGIDLSRWRELMKEEGEKADMEKFTQYVKG 544
Dc   S------M-----RT--L--T---------VQ--K-QT-GL---V-H-R- 546
EcI  KH-----C-VAN-KAL-TNVH-LN-EN-Q-ELAQAK-PFNLGRLIRL--E 539
EcII TGFEFVLA-VVD-RRS-L-YD-L-A--ALAFFND-AVEQ-E-SLFLWMRA 522

At   NHFIPNSVMVDCTADADIASCYYDWLLRGIHVVTPNKKANSGPLDQYLKI 594
Dc   -----ST-I------SEV--H-H---C-----I----------------L 596
EcI  Y-LL -P-I-N--SSQAV-DQ-A-F-RE-F----------TSSM-Y-HQL 588
EcII HPY DDL-VL-V--SQQLADQ-L-FASH-F--ISA--L-GASDSNK-RQ- 571

At   RDLQRKSYTHYFYEATVGAGLPIISTLRGLLETGDKILRIEGIFSGTLSY 644
Dc   -A---R-----------------T--Q---------------------- 646
EcI  -YAAE--RRKFL-DIN------V-EN-QN--NA--ELMKFS--L--S--- 638
EcII H-AFE-TGR-WL-N---------NH-V-D-IDS--T--S-S--------W 621

                Homoserine Dehydrogenase
At   LFNNFAGTRSFSEVVAEAKQAGFTEPDPRDDLSGTDVARKVTILARESGL 694
Dc   I----KS-TP-----S---A--Y---------A--------I-------- 696
EcI  I-GKLDEGM----ATRL-REM-Y----------M-----LL-----T-R 688
EcII --LQ-D-SVP-T-L-DQ-W-Q-L----------K--S--LV-----A-Y 671

At   KLDLEGLPIQNLVPKPLQACASAEEFMEKLPQFDEELSKQREEAEAAGEV 744
Dc   --E-SDI-V-S---E--RGI------LLQ-----SDMTRK--D--N---- 746
EcI  E-E-ADIE-EPVL-AEFN-EGDVAA--AN-S-L-DLFAARVAK-RDE-K- 738
EcII NIEPDQVRVES---AHCEG G-IDH-F-NGDELN-QMVQRL-A-REM-L- 720

At   LRYVGVVDAVEKKGT  VELKRYKKDHPFAQLSGADNIIAFTTKRYKEQP 792
Dc   ----------NQ--V  --------E--------S--------E--NK-- 794
EcI  -----NI-   ED-VCR-KIAEVDGND-LFKVKNGE-AL--YSHY-QPL- 785
EcII ----ARF--   N-KAR-GVEAVRE---LRS-LPC--VF-IESRW-RDN- 757

                                        824
At   LIVRGPGAGAQVTAGGIFSDILR LAFYLGAPS 824
Dc   --I--------E-----V------ --S------ 826
EcI  -VL--Y---ND---A-V-A-L--T-SWK--V   816
EcII -VI------RD----A-Q---N-     -AQLL 795
```

[b] Putative kinase domain

```
At    211-ALFRSHQLTIWTDVDGVYSADPRKVSEAVVLKTLSYQEAWEMSYFGANVLHPRTIIPVM
Dc    211-ALLRAGQVTIWTDVNGVYSADPRKVSEAVVLKTLSYQEAWEMSYFGANVLHPRTIIPVM
Sm    216-ACLRADCCEIWTDVDGVYTCDPRTVPDARLLKSMSYQEAMELSYFGAKVLHPRTITPIA
EcI   216-ACLRADCCEIWTDVNGVYTCDPRQVPDARLLKSMSYQEAMELSYFGAKVLHPRTITPIA
EcII  218-ALAGVSRVTIWSDVAGVYSADPRKVKDACLLPLLRLDEASELARLAAPVLHARTLQPVS
EcIII 210-EALHASRVDIWTDVPGIYTTDPRVVSAAKRIDEIAFAEAAEMATFGAKVLHPATLLPAV
BsII  159-AALKVDKCDIYTDVPGVFTTDPRYVKSARKLEGISYDEMLELANLGAGVLHPRAVEFAK
Cg    160-AALNADVCEIYSDVDGVYTADPRIVPNAQKLEKLSFEEMLELAAVGSKILVLRSVEYAR
Sc    244-VAVNADELQVWKEVDGIFTADPRKVPEARLLDSVTPEEASELTYYGSEVIHPFTMEQVI
AK cs     AAL.AD...IWTDV.GVYTADPRKV..A..L..LSY.EA.ELAYFGA.VLHPRT..PV.

AK cs           215 AD...IWTDV.GVYTADPRKV..A 238
Ec ProB         162 ADKLLLLTDQKGLYTADPRSNPQA 185
Va P5CS gene    190 ADLLVLLSDVEGLYSGPPSDPHSK 213
```

Fig. 5. a. Amino acid sequence of the *Arabidopsis thaliana* AK-HSDH and comparison from the KFGG to the termination codon of the *A. thaliana* (At) with the carrot (Dc) and the two *E. coli* AK-HSDH encoding genes, *thrA* (EcI) and *metL* (EcII). (-) and ( ) are for identical residues and gaps. b. Putative kinase domain. AK cs is the conserved (more than 50%) amino acids around the D-P-R sequence from all AK sequences available to date. Sm, *Serratia marcescens*; Ec, *Escherichia coli*, Bs, *Bacillus subtilis*, Cg, *Corynebacterium glutamicum*; Sc, *Saccharomyces cerevisiae*; Dc, *Daucus carota*; At, *A. thaliana* [14, 25–30, 32]. Ec ProB and Va P5CS gene are the D-P-R sequence respectively of *E. coli* [35] and *Vigna aconitifolia* [36].

amino acid sequence (Table 1). Indeed, sequences from position 1 to 243 and from position 503 to 824 show nearly the same value of sequence identity, 31% and 39%, to the 2 *E. coli* isozymes I

Table 1. Sequence similarities of the AK-HSDH bifunctional enzymes.

	EcI	EcII	Dc	At
E. coli AKI-HSDHI	100	29.8	37.2	37.5
E. coli AKII-HSDHII		100	31.4	30.7
D. carota AK-HSDH			100	80.3
A. thaliana AK-HSDH				100

and II. In contrast, sequences in between are much more homologous to the corresponding region of the *E. coli* isozyme I (*thrA* gene), 41% identity, than to the isozyme II (*metL* gene), with 15% identity. These three blocks match exactly the functional regions determined by proteolytic studies on the *E. coli* bifunctional enzymes [38]. Therefore, the AK and HSDH catalytic domains of the *Arabidopsis* gene are equally homologous to their *E. coli* counterparts. Only the ID region is specifically homologous to the AK I-HSDH I. The relevance of this finding is discussed below.

Polymorphism of the ak-hsdh *gene*

Sequence comparison of the λE19 and λF211 overlapping fragments, revealed only limited sequence variation. In the 1.4 kb fragment, the 37 changes were distributed as follows: 25 in the introns (5 deletions or additions and 20 substitutions) and 13 substitutions in the exons (8 silent). Thus, about 68% of the changes were localized in the introns and only 1.95% amino acid sequence divergence was observed. These are features typical of race variation. The two phage clones correspond respectively to the races Columbia and Landsberg *erecta*. Another example of the polymorphism has been observed with the *Hin*d III digestion of *Arabidopsis* total DNA.

5′- and 3′-flanking sequences

The 3 kb upstream region nucleotide sequence, starting at the KFGG consensus, was analysed for the presence of *cis*-controlling and promoter elements. Screening for CAAT box, TATA box followed by an ATG in a good context (distance and a purine in position −3) was based on Joshi's analysis of plant 5′-flanking sequences [39] and the PCGENE program EukProm [40]. These analyses identified a putative promoter, the features of which being summarized in Table 2. Interestingly, the TATA-box sequence was closer to the consensus sequences of the enzyme subclass of Joshi's analysis. Putative *cis*-controlling elements were also identified. A TGACTC motif is found upstream of the promoter elements, 170 bp from the TATA box. This finding might reflect a binding site of a plant GCN4-like transcriptional factor. Indeed, in yeast, such motifs were shown to be the binding site of GCN4, a positive regulatory factor of several enzymes of the amino acid biosynthesis pathways [41]. Sequence analysis for binding sites of known plant transcription factors reviewed by Katagiri and Chua [42] did not show any perfectly conserved binding sites. Two other binding sites of plant transcription factors were investigated. The prolamine '− 300 element', which is found in a wide range of species to be present in the 5′-flanking sequence of storage proteins encoding genes [43], is not present in the 2 kb sequence upstream of the TATA box. The *Opaque2* binding site (GATGAPyPuTGPu [44]) is found at −744 bp of the TATA box (GATGACGTGG). The *Opaque2* transcription activation factor regulates essentially the synthesis of

Table 2. Sequence analysis of the putative *ak-hsdh* promoter.

Element	Sequence	Relative position (bp)
TATA box	TATATATATAA	1
CAAT box	CCATT	− 87
	CCATC	− 47
First ATG	AATATGCCG	+ 254
Leader sequence	AT rich	+ 254 to ± 500 (200 to 250 bp in length)
GCN4-like sequence	CTTGACTCTA	− 181
Opaque2-like sequence	GATGACGTGG	− 744

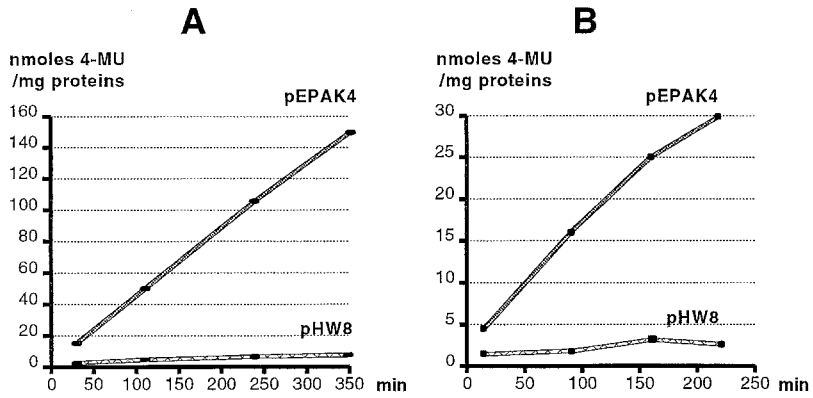


Fig. 6. Transient gene expression of the putative promoter of the *Arabidopsis thaliana ak-hsdh* gene. A. Cell suspension protoplasts (250000) were incubated each with 20 μg pEPAK4 (the 270 bp *ak-hsdh* promoter) and pHW8 (a promoter-free construct). B. Mesophyll protoplasts (500000) were treated identically.

the maize 22 kDa zeins [45] which are coordinately synthesized with enzymes in amino acid biosynthetic pathways required for their accumulation in the maize endosperm [46]. It has also been shown to regulate the synthesis of a maize 32 kDa albumin termed b-32 [47]. The involvement of this motif in the regulation of the *ak-hsdh* gene would suggest a possible common regulatory mechanism.

The 3′-flanking sequence of the *ak-hsdh* gene was analyzed for the occurrence of putative polyadenylation signals [48]. The ochre TAA codon (T is referred as position −3 in this 3′ sequence analysis), is the predominant one in higher plants with a purine in +1 but no preference for an adenine or a thymine at +3 is observed. The consensus sequence AATAAA is present twice at position +78 and +103 and a polyadenylation site is present at position +122 as evidenced by the cDNA nucleotide sequence analysis (Fig. 3). The upstream 10 nucleotides sequences flanking these motifs resemble the consensus sequence whereas the downstream 10 nt sequences do not. These AATAAA sequences, often present upstream of the polyadenylation site in plant mRNA, have been shown to have a significant effect on the formation of mRNA 3′ end [50]. Search for other less conserved sequences failed to identify such motifs.

Functional analysis of the ak-hsdh *promoter*

The sequence between the *Bgl* II site (position 1795) and the first ATG in the genomic sequence (position 2065) contains the consensus elements of the *ak-hsdh* promoter: CAAT and TATA boxes, the leader sequence and the translation initiation site analogous to the putative translation initiation site found in the cDNA sequence. This 270 bp fragment was fused to the β-glucuronidase (GUS) reporter gene so as to use the *Nco* I site as the first ATG of GUS. This construct was designated pEPAK4. Protoplasts from cell suspension cultures as well as mesophyll tissue of *N. plumbaginifolia* were used in a transient expression system with pEPAK4 and as a control pHW8, a plasmid with a promoter-free GUS gene. GUS assays show clearly that only protoplasts treated with pEPAK4 give a significant GUS activity (Fig. 6). These findings indicate that the 270 bp fragment of the *ak-hsdh* gene is most probably the functional plant promoter and thus, that the first ATG found in the cDNA sequence is the initiation codon.

Discussion

Aspartate kinase plays a major role in the biosynthesis of several essential amino acids. De-

cades of biochemical studies in higher plants did not result in a general model establishing its regulation and isozyme status. The first break-through came from the purification of another protein of the aspartate-derived amino acid pathway in carrot [13]. These authors purified the homoserine dehydrogenase and determined two internal amino acid sequences. Surprisingly, one of the sequences found was homologous to the AK region of the *E. coli thrA* gene. Association between the threonine-sensitive AK and threonine-sensitive HSDH has been suggested in pea [50]. Molecular cloning of the carrot *ak-hsdh* gene has confirmed its bifunctional structure [14].

Here, we report the molecular cloning of the analogous gene in the model plant *Arabidopsis thaliana*. Molecular study of AK in *Arabidopsis* has specific aims. In this species, a mutant resistant to growth inhibitory concentrations of lysine + threonine was shown to have an AK activity with decreased sensitivity to end-product feedback inhibition and therefore a threonine overproduction [9]. Insights gained by understanding the structural bases of the mutated alleles of AK from *Arabidopsis* may yield new approaches for improving the nutritional properties of crop plants.

The *ak-hsdh* gene has been cloned as two overlapping genomic fragments. It is interrupted by 17 introns scattered along the sequence. They are of small size as expected for *Arabidopsis* introns and are predominantly of the pyrimidine-rich class. Upstream of the KFGG conserved box of AK is a putative transit peptide sequence interrupted by one intron. The presence of a transit peptide sequence is in agreement with the localization of AK and HSDH in the chloroplast [51, 52].

The amino acid sequence from the KFGG box to the termination codon has 80.3% identity with the carrot gene. The uneven distribution of the amino acid sequence identities between the plant and *E. coli* bifunctional proteins provides evidence that the cloned plant genes are analogous to the *E. coli thrA* gene. First, the ID region is the most divergent region between the two bacterial isozymes and clearly, the homology of the plant sequence drops down to 15% with the ID *metL* gene whereas it reaches 41% with the ID *thrA* gene. Second, the ID region which is thought to be involved in subunit association and devoid of any catalytic activity is responsible for the AEC (*S*-(2-aminoethyl)-*L*-cysteine, a lysine analogue) resistance of a *Corynebacterium glutamicum* mutant [30, 31]. More recently, additional evidences that the threonine sensitivity of the AK-HSDH enzyme rely on ID, came from deletion and mutation analyses of AK-HSDH in *Serratia marcescens* [28]. Thus, the ID region bears probably the feedback regulatory site of AK which is the threonine sensitive site in the *thrA* gene product. Moreover, the carrot *ak-hsdh* gene corresponds most probably to the threonine-sensitive HSDH [13].

Although no functional expression of the plant cloned genes has been reported, our analysis indicates that the *Arabidopsis* clone probably represents the plant threonine-sensitive AK-HSDH. Moreover, some biochemical data on the maize and carrot lysine-sensitive AK show that their subunits cannot be as big as the one encoded primarily by the bifunctional gene [11, 12]. This gene is present as a single copy in *Arabidopsis* and despite relaxed hybridization conditions, no other related signals could be detected by Southern blotting analysis. The gene encoding the lysine sensitive isoform of AK is probably of low homology with the bifunctional AK-HSDH-encoding genes, as is the case in *E. coli*.

Flanking sequences were analysed for the presence of consensus sequences. At the 3′ end, a polyadenylation signal, AATAAA, was repeated twice downstream of the termination codon. The cDNA isolated ended at 13 nucleotides from the second polyadenylation signal which is in good agreement with the role of these sequences in the polyadenylation process. At the 5′ end, screening for promoter elements pointed out one putative TATA box. We made the assumption that if this promoter element was functional in the plant, the first downstream ATG should be the initiation codon according to the scanning mechanism found in eukaryotes for translation initiation. Such a codon, in a good nucleotide sequence context and at an appropriate distance, was found.

The *Bgl* II-ATG initiation codon, the 270 bp fragment, was fused to the GUS reporter gene. Functional expression of GUS was obtained with a transient expression system using protoplasts from cell suspension and mesophyll tissue of *Nicotiana plumbaginifolia*. This finding indicates that this promoter element is functional and hence, the first ATG of the cDNA sequence, the initiation codon. Interestingly, screening of the upstream promoter sequences for other conserved motifs revealed two putative regulatory elements.

1. A TGACTC sequence at 170 bp of the TATA box. This element is present in single or multiple copies at the 5′-non-coding regions of genes subject to the general amino acid control system in yeast. This putative *cis*-controlling element is a great matter of interest with regard to the regulation of the amino acid biosynthesis in higher plants.

2. An *Opaque2* regulatory element [44] is found at −744 bp of the TATA box. The *Opaque2* transactivator was reported to have a direct or indirect effect on the synthesis of enzymes of amino acid biosynthetic pathways required for zein accumulation in maize endosperm [46] and in particular on AK as shown by monitoring the amino acid overproduction of a maize *Ltr* mutant in an *Opaque2* background [53].

Although these two regulatory elements are relevant as controls of the expression of a gene coding for an enzyme limiting the carbon flux towards the biosynthesis of essential amino acids, it is not yet known whether they are functional *in vivo*.

Acknowledgements

We are thankful to Dr H. Goodman, Boston, USA, for providing us the two genomic banks, to Dr Trezzini, Max Planck Institut, Germany, for the cDNA bank and to Dr J. Botterman, Plant Genetic Systems, for the gift of plasmid pHW8. M. Vauterin is thanked for providing sequence analysis software GeneCompar. NATO is acknowledged for the NATO Collaborative Research Grant CRG 900601.

References

1. Cohen NG, Saint-Giron I: Biosynthesis of threonine, lysine and methionine. In: Neidhardt FC, Ingraham JL, Low KB, Magasanik B, Schaechter M, Umbarger HE (eds) *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology, vol I, pp. 429–444. American Society for Microbiology, Washington DC (1987).
2. Bryan JK: Advances in the biochemistry of amino acids. In: Miflin BJ, Lea PJ (eds) Intermediary Nitrogen Metabolism, pp. 161–195. The Biochemistry of Plants, vol. 16. Academic Press, New York (1990).
3. Hegel J, Frydrych Z: Optimum amino acid supplementation of a high-protein wheat. Lysine and threonine. Nutr Rep Int 36: 927–934 (1987).
4. Jacobs M, Frankard V, Ghislain M: Mutants in the biosynthesis of amino acids. In: Sangwan RS, Sangwan-Norreel BS (eds) The Impact of Biotechnology in Agriculture, pp. 247–258. Kluwer Academic Publishers, Dordrecht (1990).
5. Bright SWJ, Kueh JSH, Franklin J, Rognes SE, Miflin BJ: Two genes for threonine accumulation in barley seeds. Nature 299: 278–279 (1982).
6. Cattoir-Reynaerts A, Degryse E, Verbruggen I, Jacobs M: Selection and characterization of carrot embryoid cultures resistant to inhibition by lysine plus threonine. Biochem Physiol Pflanzen 178: 81–90 (1983).
7. Diedrick TJ, Frisch DA, Gengenbach BG: Tissue culture isolation of a second mutant locus for increased threonine accumulation in maize. Theor Appl Genet 79: 209–215 (1990).
8. Frankard V, Ghislain M, Negrutiu I, Jacobs M: High threonine producer mutant of *Nicotiana sylvestris* (Spegazzini and Comes). Theor Appl Genet 82: 273–282 (1991).
9. Vernaillen S, Van Ghelue M, Verbruggen I, Jacobs M: Characterization of mutants in *Arabidopsis thaliana* (L. Heyn.) resistant to analogues and inhibitory concentrations of amino acids derived from aspartate. Arab Inf Serv 22: 13–22 (1985).
10. Bright SWJ, Miflin BJ, Rognes SE: Threonine accumulation in the seeds of a barley mutant with an altered aspartate kinase. Biochem Genet 20: 229–243 (1982).
11. Relton JM, Bonner PLR, Wallsgrove RM, Lea PJ: Physical and kinetic properties of lysine-sensitive aspartate kinase purified from carrot cell suspension culture. Biochim Biophys Acta 953: 48–60 (1988).
12. Dotson SB, Somers DA, Gengenbach BG: Purification and characterization of lysine-sensitive aspartate kinase from maize cell cultures. Plant Physiol 91: 1602–1608 (1989).
13. Wilson BJ, Gray AC, Matthews BJ: Bifunctional protein in carrot contains both aspartokinase and homoserine dehydrogenase activities. Plant Physiol 97: 1323–1328 (1991).
14. Weisemann JM, Matthews BJ: Identification and expres-

sion of a cDNA from *Daucus carota* encoding a bifunctional aspartokinase-homoserine dehydrogenase. Plant Mol Biol 22: 301–312 (1993).
15. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual; 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).
16. Hanahan D: Studies on transformation of *Escherichia coli* with plasmids. J Mol Biol 166: 557 (1983).
17. Dellaporta SL, Wood J, Hicks JB: A plant DNA mini preparation: version II. Plant Mol Biol Rep 1: 19–21 (1983).
18. Fourney RM, MiyaKoshi J, Day III RS, Paterson MC: Northern blotting: efficient RNA staining and transfer. Focus 10: 5–7 (1987).
19. Sanger FS, Nicklen S, Coulson AR: DNA sequencing with the chain terminating inhibitors. Proc Natl Acad USA 74: 5463–5467 (1977).
20. Bilang R, Schnorf M: Mesophyll protoplasts of tobacco. In: Potrykus I (ed) Gene Transfer to Plants, pp. 18–25. EMBO Advanced Laboratory Course, Swiss Federal Institut of Technology (1991).
21. Jefferson RA, Kavanagh TA, Bevan MW: GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J 6: 3901–3907 (1987).
22. Keegstra K, Olsen LJ, Theg SM: Chloroplastic precursors and their transport across the envelope membranes. Annu Rev Plant Physiol Plant Mol Biol 40: 471–501 (1989).
23. Brown JWS: A catalogue of splice junction and putative branch point sequences from plant introns. Nucl Acids Res 14: 9549–9559 (1986).
24. Kaneko T, Hashimoto T, Kumpaisal R, Yamada Y: Molecular cloning of wheat dihydrodipicolinate synthase. J Biol Chem 265: 17451–17455 (1990).
25. Cossart P, Katinka M, Yaniv M, Saint-Girons I, Cohen GN: Nucleotide sequence of the *thr* B gene of *E. coli*, and its two adjacent regions: the *thr* AB and *thr* C junctions. Nucl Acids Res 9: 339–347 (1981).
26. Zakin MM, Duchange N, Ferrara P, Cohen GN: Nucleotide sequence of the *met* L gene of *Escherichia coli*. Its product, the bifunctional aspartokinase II-homoserine dehydrogenase II and the bifunctional product of the *thr* A gene, aspartokinase I-homoserine dehydrogenase I derive from a common ancestor. J Biol Chem 258: 3028–3031 (1983).
27. Cassan M, Parsot C, Cohen GN, Patte JC: Nucleotide sequence of the *lysC* gene encoding the lysine sensitive aspartokinase III of *Escherichia coli* K12: evolutionary pathway leading to three isofunctional enzymes. J Biol Chem 261: 1052–1057 (1986).
28. Omori K, Imai Y, Suzuki S-I, Komatsubara S: Nucleotide sequence of the *Serratia marcescens* threonine operon and analysis of the threonine operon mutations which alter feedback inhibition of both aspartokinase I and homoserine dehydrogenase I. J Bact 175: 785–794 (1993).
29. Chen NY, Hu FM, Paulus H: Nucleotide sequence of the overlapping genes for the subunits of *Bacillus subtilis* aspartokinase II and their control regions. J Biol Chem 262: 8787–8798 (1987).
30. Kalinowski J, Bachmann B, Thierbach G, Pühler A: Aspartokinase genes *lysCa* and *lysCb* overlap and are adjacent to the aspartate β-semialdehyde dehydrogenase gene *asd* in *Corynebacterium glutamicum*. Mol Gen Genet 224: 317–324 (1990).
31. Kalinowski J, Cremer J, Bachmann B, Eggeling L, Sahm H, Pühler A: Genetic and biochemical analysis of the aspartokinase from *Corynebacterium glutamicum*. Mol Microbiol 5: 1197–1204 (1991).
32. Rafalski AJ, Falco CS: Structure of the yeast *hom3* gene which encodes for aspartokinase. J Biol Chem 263: 2146–2151 (1988).
33. Hanley BA, Schuler MA: Plant intron sequences: evidence for distinct groups of introns. Nucl Acids Res 16: 7159–7175 (1988).
34. Omori K, Suzuki S-I, Imai Y, Komatsubara S: Analysis of the *pro*BA operon and feedback control of proline biosynthesis. J Gen Microbiol 137: 509–517 (1991).
35. Deutch AH, Rushlow KE, Smith CJ: Analysis of the *Escherichia coli pro*BA locus by DNA and protein sequencing. Nucl Acids Res 12: 6337–6355 (1984).
36. Hu C-A, Delauney AJ, Verma DPS: A bifunctional enzyme (D1-pyrroline-5-carboxylate synthetase) catalyses the first two steps in proline biosynthesis in plants. Proc Natl Acad Sci USA 89: 9354–9358 (1992).
37. Campbell WH, Gowri G: Codon usage in higher plants, green algae and cyanobacteria. Plant Physiol 92: 1–11 (1990).
38. Fazel A, Müller K, Le Bras G, Garel J-R, Veron M, Cohen GN: A triglobular model for the polypeptide chain of aspartokinase I-homoserine dehydrogenase I of *Escherichia coli*. Biochemistry 22: 158–165 (1983).
39. Joshi CP: An inspection of the domain between putative TATA box and translational start site in 79 plant genes. Nucl Acids Res 15: 6643–6653 (1987a).
40. Bucher P: Wheight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. J Mol Biol 212: 563–578 (1990).
41. Arndt K, Fink GR: GCN4 protein, a positive transcription factor in yeast, binds general control promoters at all 5′ TGACTC 3′ sequences. Proc Natl Acad Sci USA 83: 8516–8520 (1986).
42. Katagiri K, Chua N-H: Plant transcription factors: present knowledge and future challenges. Trends Genet 8; 22–27 (1992).
43. Forde BG, Heyworth A, Pywell J, Kreis M: Nucleotide sequence of a Bl hordein gene and the identification of possible upstream regulatory elements in endosperm storage protein genes from barley, wheat and maize. Nucl Acids Res 13: 7327–7339 (1985).
44. Lohmer S, Maddaloni M, Motto M, Di Fonzo N,

Hartings H, Salamini F, Thompson RD: The maize regulatory locus *Opaque-2* encodes a DNA-binding protein which activates the transcription of the b-32 gene. EMBO J 10: 617–624 (1991).

45. Kodrzychi R, Boston RS, Larkins BA: The *opaque 2* mutation of maize differentially reduces zein gene transcription. Plant Cell 1: 105–114 (1989).
46. Motto M, Di Fonzo N, Hartings H, Maddaloni M, Salamini F, Soave C, Thompson RD: Regulatory genes affecting maize storage protein synthesis. Oxford Surv Plant Mol Cell Biol 6: 87–114 (1989).
47. Bass HW, Webster C, OBrian GR, Roberts JKM, Boston RS: A maize ribosome-inactivating protein is controlled by the transcriptional activator Opaque 2. Plant Cell 4: 225–234 (1992).
48. Joshi CP: Putative polyadenylation signals in nuclear genes of higher plants: a compilation analysis. Nucl Acids Res 15: 9627–9640 (1987b).
49. Mogen BD, MacDonald MH, Graybosh R, Hunt AG: Upstream sequences other than AAUAAA are required for efficient messenger RNA 3′-end formation in plants. Plant Cell 2: 1261–1272 (1990).
50. Aarnes H, Rognes SE: Threonine sensitive aspartate kinase and homoserine dehydrogenase from *Pisum sativum*. Phytochemistry 13: 2717–2724 (1974).
51. Bryan JK, Lissik EA, Matthews BF: Changes in enzyme regulation during growth of maize. III. Intracellular localization of homoserine dehydrogenase in chloroplasts. Plant Physiol 59: 673–679 (1977).
52. Wallsgrove RM, Lea PJ, Miflin BJ: Intracellular localization of aspartate kinase and the enzymes of threonine and methionine biosynthesis in green leaves. Plant Physiol 71: 780–784 (1983).
53. Azevedo RA, Arana JL, Arruda P: Biochemical genetics of the interaction of the lysine plus threonine resistant mutant *Ltr*1* with opaque-2 maize mutant. Plant Sci 70: 81–90 (1990).